International Journal of Engineering Research and Applications *www.ijera.com*
ISSN: 2248-9622, pp. 20-22

RESEARCH ARTICLE OPEN ACCESS

A Review on Data Science and its Technologies, life Cycle and Application

Abhay Pandey*, Neha Singh**
**Student, Department of Computer Science and Engineering,Raj Kumar Goel Institute of Technology, India*
***Asst. Professor, Department of Computer Science and EngineeringRaj Kumar Goel Institute of Technology, India*

ABSTRACT
We are living in the 21 CE where daily the lot of data is generated in daily life. Older method and technique were not very efficient to handle. Then the technology come known data science. Data science is one of the most secure and safe way of managing data which is generated day by day. The main goal of the data science is to convert raw material into useful way. The information in this paper will be what is data science? What is the use and need of data science, what methods and programming languages are used in data science. How many different kind of libraries are used in finding the solution of problems. Some basic mathematics is also used in data science.
***Keywords*- python, R, Big Data, Hadoop, tableau, Data Analysis**

I. INTRODUCTION

Data Science [5] [9] is a best combination of the different categories like arithmetic, measurements python or R programming Language. There are languages are also used for making the data valuable to increase the demand of the company. also some quick methods which are used for catching information that may not be caught at this moment in addition to the capacity to take a gander at things in an unexpected way. Now it is the time to think about of the extent of details and the surge of data and information which is generated daily and is provided nowadays through the advancements in emerging technologies and the internet. With the huge increase in storage capabilities and different methods of data collection, huge amounts of data have become easily available. In Every second, the huge and huge amount of data is being created and needs to be stored and analysed in order to extract valuable and useful data. Now these days' data has become cheaper to store, so different organizations now only need to get as much value as possible from the huge amounts of stored data.

Let's understand by this fig. 1 that what actually data science is. Data Science [10] means the combination of the different types of task to handle data the task which we use are programming, domain knowledge, basic maths and algorithms.

The data is collected from different areas. Then the data is analysing, clean and process. Programming languages are also used for making the data valuable to increase the demand of the company.

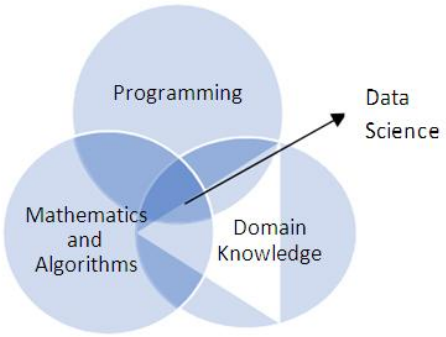

Fig. 1 Structure of Data science

As in our daily life the data is increasing the day by day so how can data science [4] can help in the managing of the data. What are the different types of tool use in data science? What will be the impact of the data science with the relevant data? There are different definition of data science as "*data science* is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data" by Provost and Fawcett, 2013 .There are many definition but according to me this is complete definition because in other words we can say this data science is a multidisciplinary field that deals with the processing and analysis of the raw data to make the data useful and then we can use various methods on data.

II. TOOLS OF DATA SCIENCE TECHNOLOGIES

Now it is the time to know the different types of tools [7] that are used for Data Science as show in fig 2.

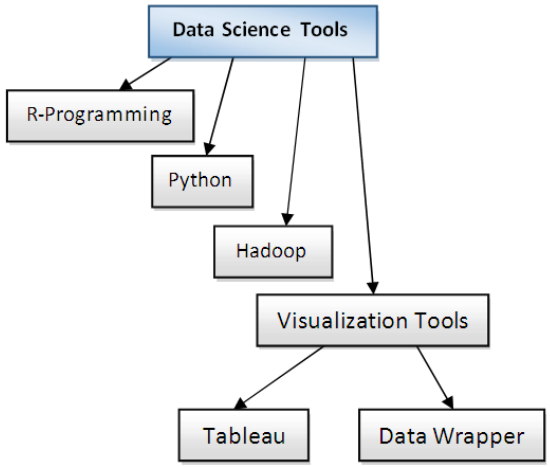

Fig. 2 Technologies used in Data Science

2.1.R-Programming

R is a programming language [2] that is widely used for statistical computing and graphics which is supported by the R Foundation for Statistical Computing. The R Programming language is now widely used by various of the statisticians and data miners for the development of the various statistical software and data analysis. R Language was created by Ross Ihaka and Robert Gentleman at the University Of Auckland, New Zealand, and it was currently developed by the R Development Core Team. After the development of R-language it's reputation has started extended impressively which was exhibited by the further studies and surveys.

2.2 Python

Python [1] is one of the most general purpose programming languages. Python has very easy syntax to learn and use this makes a python most amazing programming language to learn. It is interpreted and high level programming language. It was developed by Guido Van Rossum. Python is just a straightforward sentence structure is exceptionally available to programming .There are many library are used for the analysis of the data. For ex- matplotlib is the library of the python which is used for the creation and analysis of the graph. It can be any kind of the graph. Another library is numpy which is used for the mathematical calculations.

2.3 Hadoop

The term Apache Hadoop [3] is a Java based open source software framework which is basically used for storing huge data and running applications on clusters of hardware and using a network of many computers to solve the problems of involving massive and huge amounts of data and computation. It also provides a software framework which is used for distributed storage and processing of big data using the MapReduce programming model.

2.4 Visualization Tools

Data visualization [5] is one of the most advanced tools for the distinct measurements. It includes various tasks like the creation and investigation of the visual portrayal of information, signifying today the data that has been dreamy in some schematic frame, including properties for the units of data. Some of the tools are like:

- Tableau

Tableau Software has now become the most interactive way of data visualization. Tableau is an analytics platform so we use the different type of data to solve the problem and make the data for ouruseful task.

- Data Wrapper

Data wrapper is a technology is used for outlines and map in four stages. It can be anything but it is difficult to use we should simply transform our information and use an outline and distribute it.

III. DATA SCIENCE LIFE CYCLE

Basically the data science is not a single cyclic process [4] it is divided into different phases of life cycle as show in fig. 3. Data Science Life Cycle is consist of five important phases.

- First phase is consist of *Data Discovery*, which involves retrieving data from various sources like daily how many people are doing retweet or doing like the photos on facebook. This data will be in unstructured format. Unstructured format of data is format that is not uniform and standardised. The most common type of unstructured data is text from documents, transcripts, social media sites, etc. Other types of unstructured data are generated by smartphones like videos, images, audio files etc.
- Second phase is consisting of the *Data Preparation* in which data is collected from different-different format snow transformed into a particular specific format. It includes basic task like the cleaning of the data, reduction of the data and the transformation of the data. There are various tasks like the building of the mathematical models also.

- In third phase some different *statically formula and visualization tools* are also used to understand the relation of learning technique to analyse the model.
- Fourth phase is consisting of *operation on data* useful information is collected and possible outcomes are making according to our need.
- Last and fifth phase is *consisting of useful result.*

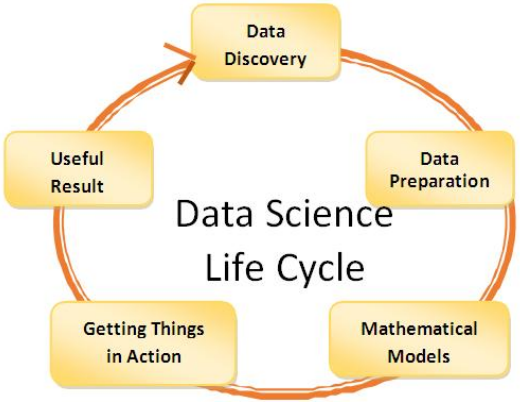

Fig.3 Life Cycle of Data Science

IV. APPLICATION OF DATA SCIENCE

Application of data science [5] [6] [8] is huge and huge. There are lot of sector in which the application of data science is used some common sector are like banking, medicine, business and transportation for solving the problem of banking fraud detection, online advertisement, marketing and inter related selling. In the field of the medicine is playing a most important role. Today many multinational companies are using data in business intelligence to increase the business of their company and this technology has improved them lot in business. Technique of data science is also used by online shopping companies to make good and better understanding with the customer so that what the exactly is looking for or what the customer need.

V. CONCLUSION

Today With the huge and enormous increase in data day by day, there are a various types of needs for analysing a large amount of data. Data Science is able manage this data and develop many types of machine learning models and algorithms that can predict the better future result in coming years. Today data science has various applications in many fields.

REFERENCE

[1]. Python, https://www.google.com/search?q=Python+progra mming.
[2]. R-ptogramming language, en.wikipedia.org › wiki › R_(programming_language).
[3]. Hadoop, hadoop.apache.org.
[4]. Dataflair Team. (2019). What is data science? : a complete data science tutorial for beginners [Blog]. Retrieved 8. 10. 2019 from https://data- flair.training/blogs/what-is-datascience.
[5]. R. Jayantilal JaiswalA Review on Data Science TechnologiesInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology| Volume 3 | Issue 3 | ISSN : 2456-3307.
[6]. Genetics Home Reference (2018) "What Is Precision Medicine?" Genetics Home Reference. Accessed May 8, 2018. https://ghr.nlm.nih.gov/primer/precisionmedicine/ definition.
[7]. Chaturapruek S, Dee T, Johari R, et al. How a data-driven course planning tool affects college students' GPA: evidence from two field experiments[C], Proceedings of the Fifth Annual ACM Conference on Learning at Scale, No. 63. London, United Kingdom, June 26-28, 2018.
[8]. Saunders, M. N., Lewis, P., & Thornhill, A. (2019). Research methods for business students. New York: Pearson.
[9]. What is a data scientist [EB/OL], https://datascopeanalytics.com/blog/ what-is-a-data-scientist.
[10]. Barber, M. (2018). Data science concepts you need to know! Part 1.Retrieved 15. 9. 2019.